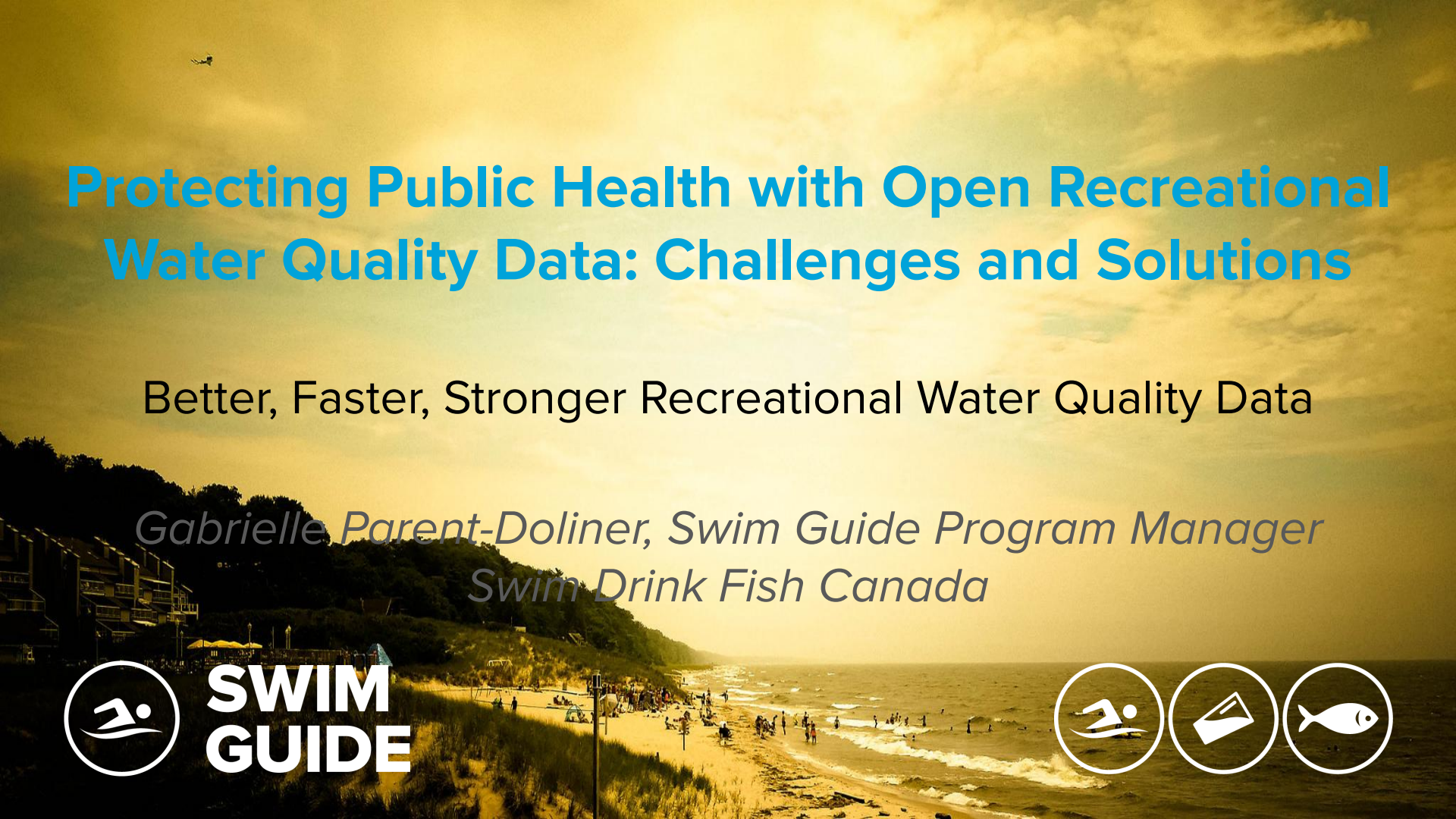
Protecting Public Health with Open Recreational Water Quality Data: Challenges and Solutions

Better, Faster, Stronger Recreational Water Quality Data

Gabrielle Parent-Doliner, Swim Guide Program Manager
Swim Drink Fish Canada



SWIM DRINK FISH

Connecting people with water

LAKE ONTARIO WATERKEEPER®

SWIM GUIDE

WATERMARK PROJECT

GREAT LAKES CHALLENGE

Great Lakes Guide

What we'll cover today

- What is open data?
- What is an open data standard?
- Swim Drink Fish open data standard
- Open recreational water quality data tips

What is Open Data?

Open data is defined as structured data that is machine-readable, freely shared, used and built on without restrictions.
(Gov of Canada - Open Data).

Open data's value is its ability to be shared.
The goal is interoperability.

The Common Pot

Sharing data is rooted in the scientific community. Researchers and librarians have known and championed the benefits of open, shared knowledge for centuries. Information technologies gave new breath to this philosophy of commons.

Transparency, participation and collaboration fosters open knowledge.

Merton: Each researcher must contribute to the “common pot” and give up intellectual property rights to allow knowledge to move forward. (1942)

Open Data in US and Canada

Both US and Canadian Federal Governments made open and machine readable the new default.

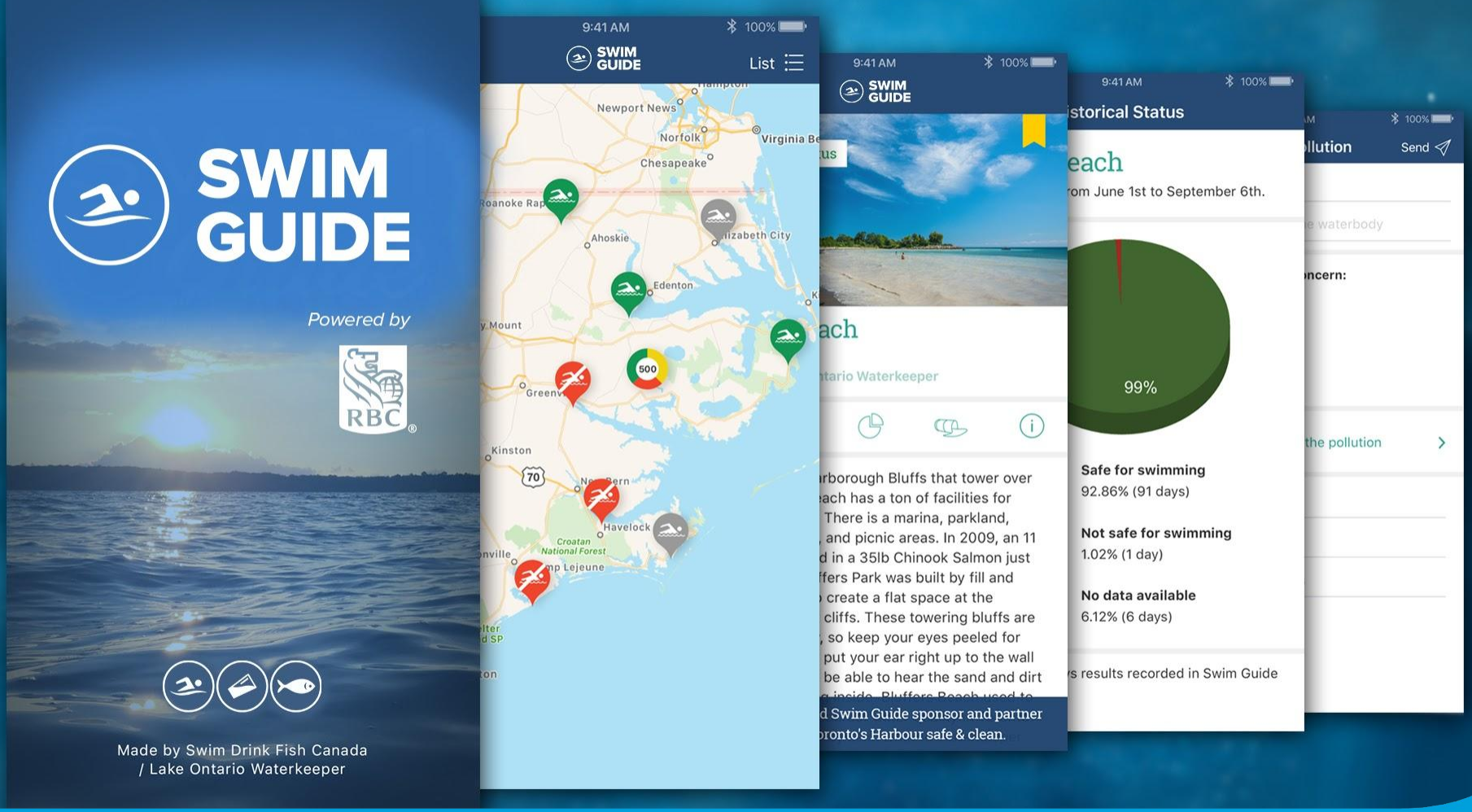
The Open Definition gives full details on the requirements for ‘open’ data and content. Open data are the building blocks of open knowledge. Open knowledge is what open data becomes when it’s useful, usable and used.

The key features of openness are:

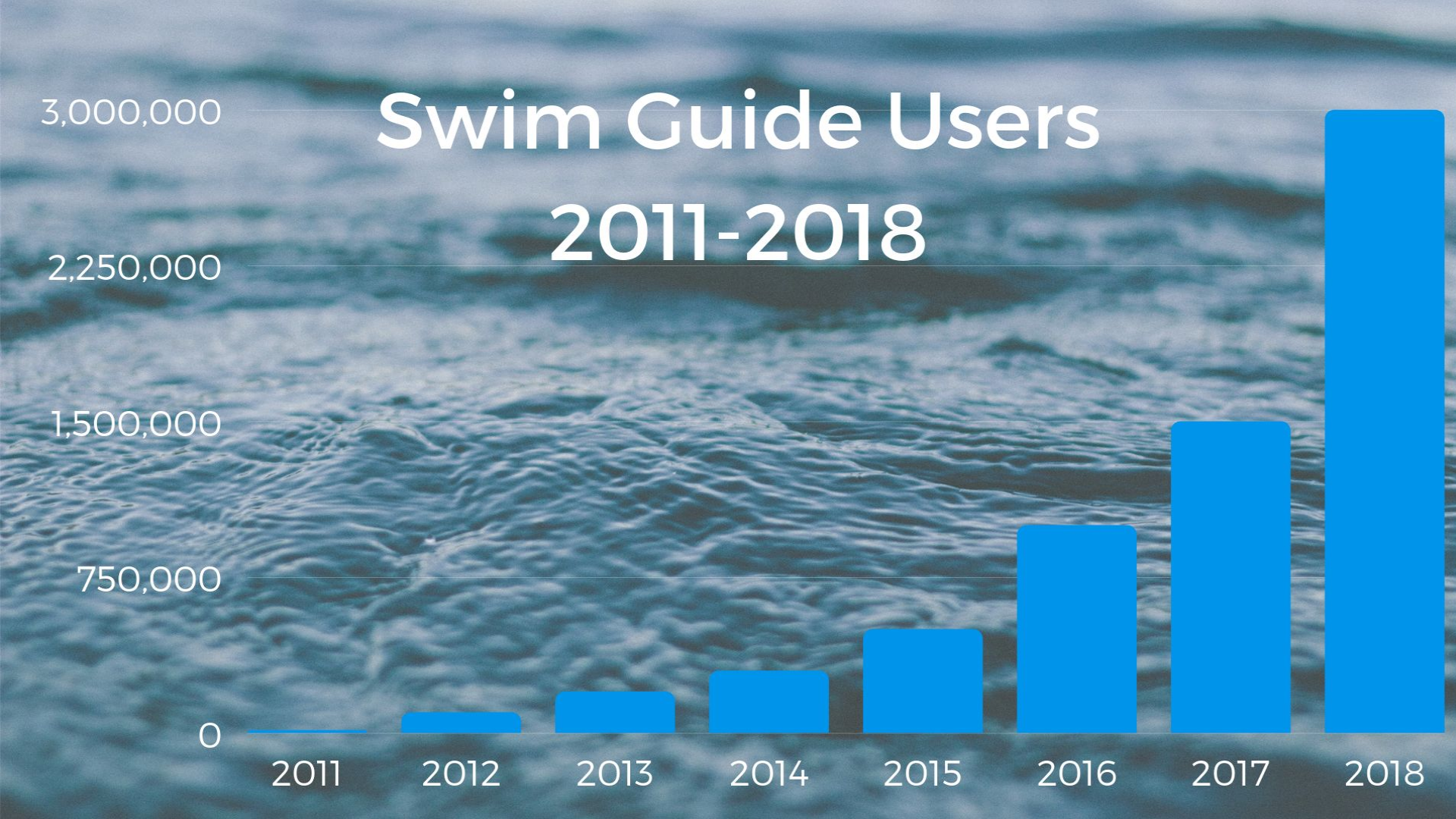
- **Availability and access**: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.
- **Universal participation:** everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

-- Open Knowledge International


Better, faster answers to the question
“ Where Can I Swim ?”
with interoperable data


SWIM GUIDE
Powered by
RBC
Made by Swim Drink Fish Canada / Lake Ontario Waterkeeper
9:41 AM
100%
SWIM GUIDE
List
Newport News
Norfolk
Virginia Be
Chesapeake
Roanoke Rap
Ahoskie
Edenton
Greenv
Kinston
70
Havelock
Croatan National Forest
500
9:41 AM
100%
SWIM GUIDE
ach
ntario Waterkeeper
rborough Bluffs that tower over
each has a ton of facilities for
There is a marina, parkland,
and picnic areas. In 2009, an 11
d in a 35lb Chinook Salmon just
ffers Park was built by fill and
create a flat space at the
cliffs. These towering bluffs are
so keep your eyes peeled for
put your ear right up to the wall
be able to hear the sand and dirt
d Swim Guide sponsor and partner
ronto's Harbour safe & clean.
9:41 AM
100%
istorical Status
each
rom June 1st to September 6th.
99%
Safe for swimming
92.86% (91 days)
Not safe for swimming
1.02% (1 day)
No data available
6.12% (6 days)
s results recorded in Swim Guide
100%
llution
Send
e waterbody
ncern:
the pollution

Swim Guide Users 2011-2018



Community and Government Water Quality Monitoring Programs Data

OPEN DATA

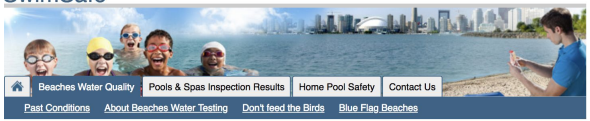
- Raw
- Machine readable/structured data (CSV, XML, JSON)
- Discoverable (open data portal)
- Re-usable, legally and technically

NOT OPEN DATA

- Processed/refined data
- Human readable/unstructured data (Word docs, PDFs, scans, pictures)
- Not findable/hidden
- Finite/restricted use

Human readable, unstructured data

SwimSafe



A - Marie Curtis Park East Beach

The Foot of 42nd St., South of Lakeshore Rd. W.

Sampled on:	September 3, 2018
Posted on:	September 4, 2018
E. coli count:	38
Advisory:	Beach meets provincially established safety standards for swimming.

Conditions are based upon E. coli counts in beach water samples taken over the past twenty-four hours.

Reporting is closed until June 2019

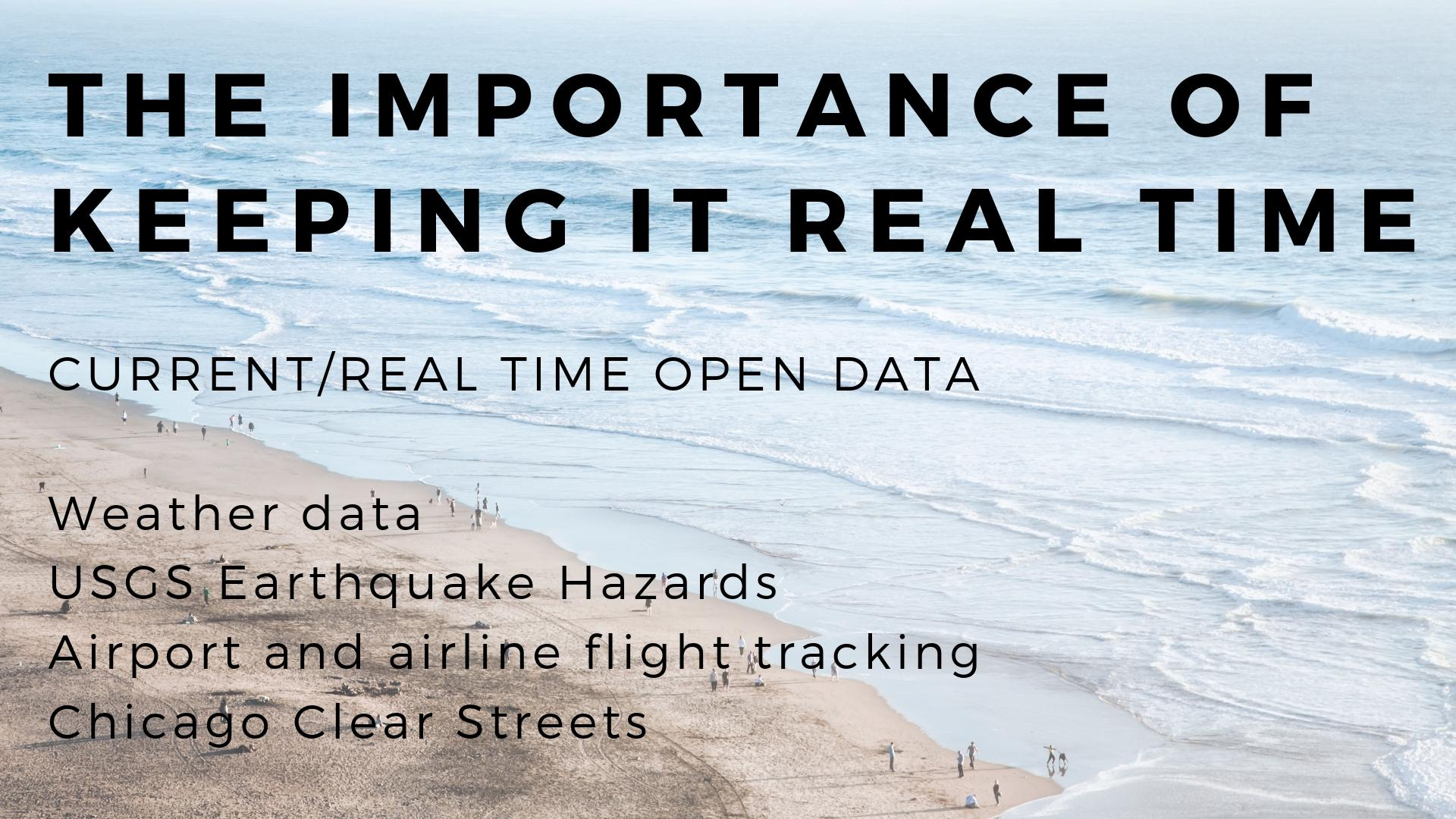
Swimming conditions history: 2018

Sample Date	E. coli Level	Swimming Condition
September 3, 2018	38	Beach meets provincially established safety standards for swimming.
September 2, 2018	--	Improving weather conditions and the local forecast indicate a decreasing trend in E. coli levels.
September 1, 2018	--	Improving weather conditions and the local forecast indicate a decreasing trend in E. coli levels.
August 31, 2018	61	Beach meets provincially established safety standards for swimming.

Open, structured data

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<Results>
  ▼<Row>
      <BEACH_ID>GA781891</BEACH_ID>
      <BEACH_NAME>LITTLE CUMBERLAND</BEACH_NAME>
      <BEACH_DESCRIPTION>Barrier island beach</BEACH_DESCRIPTION>
      <STATION_CODE>LCUM</STATION_CODE>
      <AREA_CODE>CMB</AREA_CODE>
      <LATITUDE_MEASURE>30.97782</LATITUDE_MEASURE>
      <LONGITUDE_MEASURE>-81.41448</LONGITUDE_MEASURE>
      <DATE_COLLECTED>11/01/2004</DATE_COLLECTED>
      <ENTERO>1</ENTERO>
      <ENTERO_DATE>11/02/2004</ENTERO_DATE>
      <STATUS>Gray</STATUS>
  </Row>
  ▼<Row>
      <BEACH_ID>GA543512</BEACH_ID>
      <BEACH_NAME>CUMBERLAND</BEACH_NAME>
      <BEACH_DESCRIPTION>Barrier island beach</BEACH_DESCRIPTION>
      <STATION_CODE>CUM</STATION_CODE>
      <AREA_CODE>CMB</AREA_CODE>
      <LATITUDE_MEASURE>30.7639</LATITUDE_MEASURE>
      <LONGITUDE_MEASURE>-81.45827</LONGITUDE_MEASURE>
      <DATE_COLLECTED>11/03/2004</DATE_COLLECTED>
      <ENTERO>10</ENTERO>
      <ENTERO_DATE>11/04/2004</ENTERO_DATE>
      <STATUS>Gray</STATUS>
  </Row>
  ▼<Row>
      <BEACH_ID>GA364044</BEACH_ID>
      <BEACH_NAME>CABRETTA (SAPELO)</BEACH_NAME>
      <BEACH_DESCRIPTION>Barrier island beach</BEACH_DESCRIPTION>
      <STATION_CODE>SAPC</STATION_CODE>
      <AREA_CODE>SAP</AREA_CODE>
      <LATITUDE_MEASURE>31.4325</LATITUDE_MEASURE>
      <LONGITUDE_MEASURE>-81.23883</LONGITUDE_MEASURE>
      <DATE_COLLECTED>11/22/2004</DATE_COLLECTED>
      <ENTERO>34</ENTERO>
      <ENTERO_DATE>11/23/2004</ENTERO_DATE>
      <STATUS>Gray</STATUS>
  </Row>
```



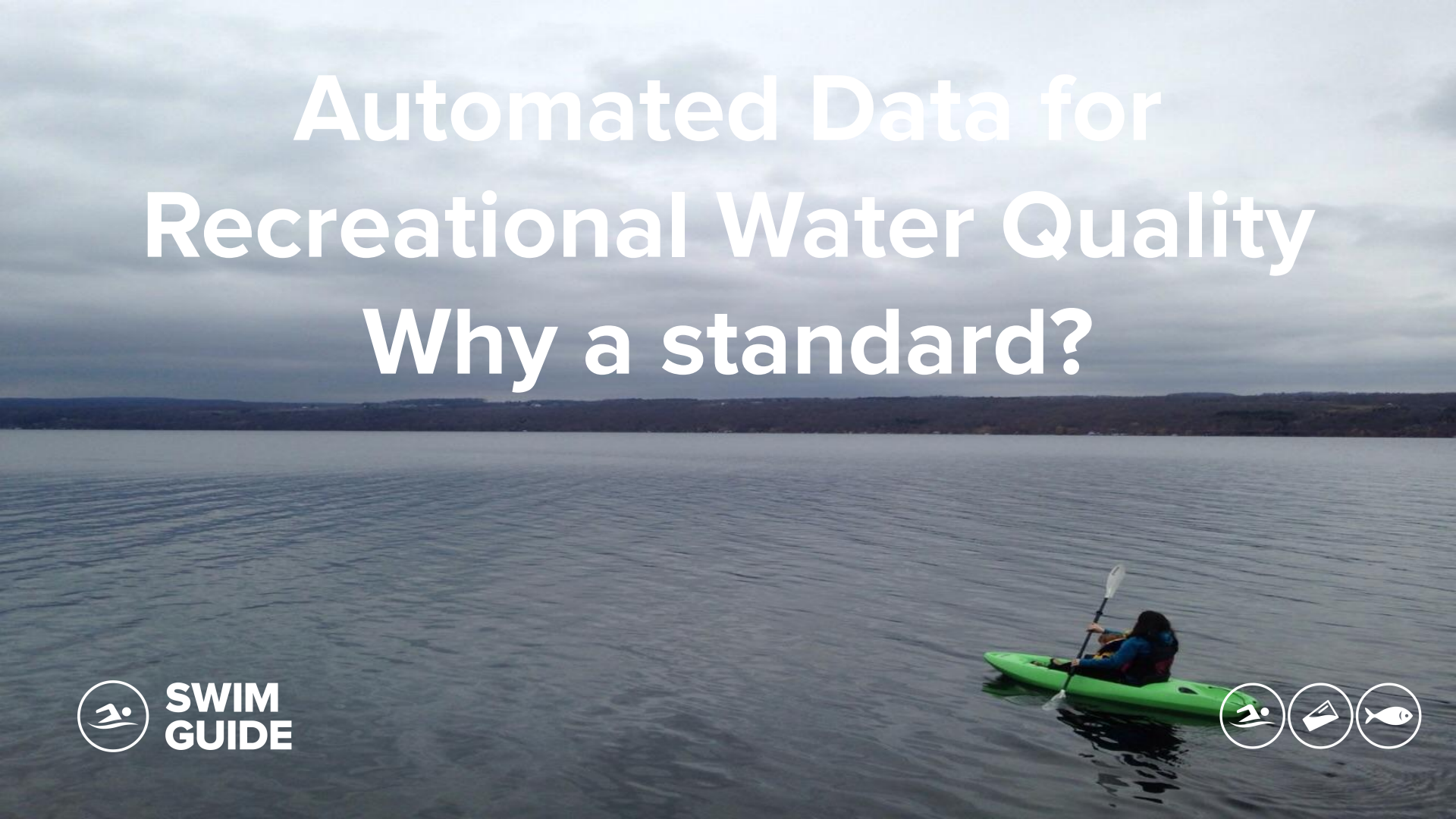
THE IMPORTANCE OF KEEPING IT REAL TIME

CURRENT/REAL TIME OPEN DATA

Weather data
USGS Earthquake Hazards
Airport and airline flight tracking
Chicago Clear Streets

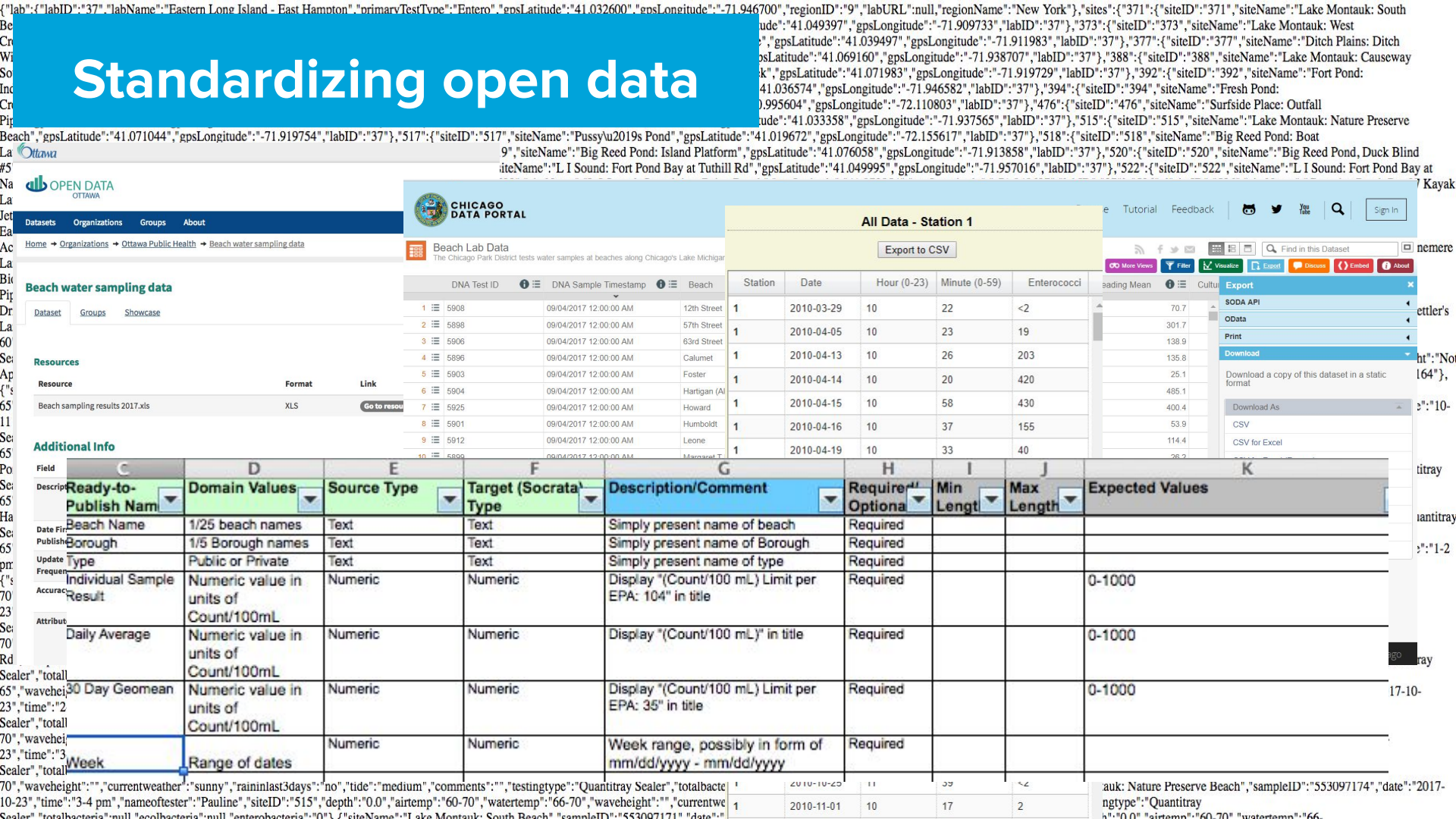
Automated Data for Recreational Water Quality

Why a standard?



SWIM GUIDE

Standardizing open data



C	D	E	F	G	H	I	J	K
Ready-to-Publish Nam	Domain Values	Source Type	Target (Socrata\ Type	Description/Comment	Require/ Optiona	Min Lengt	Max Length	Expected Values
Beach Name	1/25 beach names	Text	Text	Simply present name of beach	Required			
Borough	1/5 Borough names	Text	Text	Simply present name of Borough	Required			
Type	Public or Private	Text	Text	Simply present name of type	Required			
Individual Sample Result	Numeric value in units of Count/100mL	Numeric	Numeric	Display "(Count/100 mL) Limit per EPA: 104" in title	Required			0-1000
Daily Average	Numeric value in units of Count/100mL	Numeric	Numeric	Display "(Count/100 mL)" in title	Required			0-1000
30 Day Geomean	Numeric value in units of Count/100mL	Numeric	Numeric	Display "(Count/100 mL) Limit per EPA: 35" in title	Required			0-1000
Week	Range of dates	Numeric	Numeric	Week range, possibly in form of mm/dd/yyyy - mm/dd/yyyy	Required			

FAIR Data Principles

Open Data Standards are meant to render data:

- Findable
- Accessible
- Interoperable
- Reusable

Schema : Data structure. A schema defines elements that a file may contain. It provides a specific structure for open data (XML, JSON, CSV)

Metadata : Data about data. Metadata is a description and context of the data, such as fields or column headers. Helps to organize, find and understand data.

Data: information

A	B	C	D	E	F	G	H	I	J	K	L	M	N
locationId	locationName	latitude	longitude	sampleId	collectionDate	collectionTime	method	substance	units	result	advisory	collector	sampleKind
6A	Marina_Four	43.638809	-79.3843384	ID18 - 6A-38	2018-09-27	9:52:00 AM	9223B_colilert	ecoli	MPN	146.7	no		single
6B	Marina_Four	43.638809	-79.3843384	ID18 - 6A-38	2018-09-27	9:55:00 AM	9223B_colilert	ecoli	MPN	127.4	no		single
6C	Marina_Four	43.6384066	-79.3843933	ID18 - 6B -38	2018-09-27	10:00:00 AM	9223B_colilert	ecoli	MPN	44.1	no		single
6D	Marina_Four	43.6380505	-79.3843397	ID18 - 6C-38	2018-09-27	10:07:00 AM	9223B_colilert	ecoli	MPN	60.2	no		single
6D-DUP	Marina_Four	43.6380505	-79.3843397	ID18 - 6D-38-DUP	2018-09-27	10:10:00 AM	9223B_colilert	ecoli	MPN	57.3	no		single
6E	Marina_Four	43.6377589	-79.3843553	ID18-6E-38	2018-09-27	10:17:00 AM	9223B_colilert	ecoli	MPN	73.3	no		single
6F	Marina_Four	43.638809	-79.3843384	ID18-6F-38	2018-09-27	9:46:00 AM	9223B_colilert	ecoli	MPN	119.8	no		single
6	Marina_Four	43.6377402	-79.3849471	ID18- 6GMT-38	2018-09-27	10:00:55 AM	9223B_colilert	ecoli	MPN	77.81323819	no		geomean

Open Data Standard for Recreational Water Quality : Development Timeline

Oct 2017 : Expert workshop to lay the groundwork for the open data standard

Nov 2017: Draft consultation period begins

Feb 2018: Comment period ends

March 2018: Pilot projects begins

June 2018: Launch of Open Data Standard

Open data standard for the automated exchange of rec water quality data

JSON.Example Version 1
Github: swimdrinkfish/opendata

https://github.com/swimdrinkfish/opendata/tree/master/v1.0

simplepanda Made advisory state clearer (fixed spelling, added for example in fin... 5d4295d on Apr 5

1 contributor

58 lines (58 sloc) | 1.68 KB Raw Blame History


```
{
    "$schema": "https://raw.githubusercontent.com/swimdrinkfish/opendata/master/v1.0/schema.json#",
    "documentTime": "2018-04-04T10:00:00-05:00",
    "records": [
    {
        "guid": "ca.waterkeeper/8099-1-u3723",
        "publicationTime": "2018-04-02T08:00:00-05:00",
        "updateTime": "2018-04-03T08:00:00-05:00",
        "organizationName": "Lake Ontario Waterkeeper (http://www.waterkeeper.ca)",
              "references": [
              {
                  "guid": "ca.waterkeeper/8099-1-u3721"
              },
              {
                  "guid": "ca.waterkeeper/8099-1-u3722"
              }
          ],
        "advisory":
        {
            "issued": true,
               "description": "This advisory has been issued as this sample has exceeded the Ontario recreational water qualit
        },
        "location":
        {
            "id": "8099",
            "name": "The Main Public Beach",
            "coordinate":
            {
                "latitude": 43.636436,
                "longitude": -79.396927
            }
        },
        "sample":
        {
            "collectionTime": "2018-03-31T16:00:00-05:00",
            "location":
            {
                "id": "8099-1",
                "name": "The Main Public Beach - Location #1",
                "coordinate":
                {
                    "latitude": 43.636436,
                    "longitude": -79.396927
                }
            },
            "type":
            {
                "kind": "geomean",
                "variant": "temporal",
                "hours": 1
            },
            "substance": "ecoli",
            "method": "9223B_colilert",
            "units": "cfu",
            "result": 299.7
        }
    }]
}
```

Challenges to opening data:

Lack of good data management tools.
In-house capacity and technological limitations
Adapting current data management system to new model
Funding: lack of or funding restrictions

What can you do today? 3 easy steps

1. Ensure all data is online, not on paper or computer.
2. Clean up your spreadsheets: organized, clean, consistent
3. Share : accessible, machine readable

Tabular dataset, organized into rows and columns, and downloadable as a CSV

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
locationId	locationName	latitude	longitude	sampleId	collectionDate	collectionTime	method	substance	units	result	advisory	collector	sampleKind	sampleVariant	hours	reference
6A	Marina_Four	43.638809	-79.3843384	ID18 - 6A-38	2018-09-27	9:52:00 AM	9223B_colilert	ecoli	MPN	146.7	no		single			ID18- 6GMT
6B	Marina_Four	43.638809	-79.3843384	ID18 - 6A-38	2018-09-27	9:55:00 AM	9223B_colilert	ecoli	MPN	127.4	no		single			ID18- 6GMT
6C	Marina_Four	43.6384066	-79.3843933	ID18 - 6B -38	2018-09-27	10:00:00 AM	9223B_colilert	ecoli	MPN	44.1	no		single			ID18- 6GMT
6D	Marina_Four	43.6380505	-79.3843397	ID18 - 6C-38	2018-09-27	10:07:00 AM	9223B_colilert	ecoli	MPN	60.2	no		single			ID18- 6GMT
6D-DUP	Marina_Four	43.6380505	-79.3843397	ID18 - 6D-38-DUP	2018-09-27	10:10:00 AM	9223B_colilert	ecoli	MPN	57.3	no		single			ID18- 6GMT
6E	Marina_Four	43.6377589	-79.3843553	ID18-6E-38	2018-09-27	10:17:00 AM	9223B_colilert	ecoli	MPN	73.3	no		single			ID18- 6GMT
6F	Marina_Four	43.638809	-79.3843384	ID18-6F-38	2018-09-27	9:46:00 AM	9223B_colilert	ecoli	MPN	119.8	no		single			ID18- 6GMT
6	Marina_Four	43.6377402	-79.3849471	ID18- 6GMT-38	2018-09-27	10:00:55 AM	9223B_colilert	ecoli	MPN	77.81323819	no		geomean	spatial		

Unstructured sheet, not organized into rows and columns. Cannot be downloaded as a “usable” CSV

eeeee

Maxxam Job #: B3B2525

Report Date: 2013/07/14

Lake Ontario Wate

Client Project #:

Site Location:

Sampler Initials:

MICROBIOLOGY (WATER)

Maxxam ID		SF9121	SF9122	SF9123	SF9124	SF9125	SF9126	SF9127	SF9128	SF9129	SF9130	SF9131
Sampling Date		7/12/13 11:28	7/12/13 11:37	7/12/13 11:39	7/12/13 12:39	7/12/13 12:38	7/12/13 13:10	7/12/13 13:13	7/12/13 13:20	7/12/13 14:55	7/12/13 14:55	7/12/13 14
COC Number		25361	25361	25361	25361	25361	25361	25361	25361	25361	25361	25361
	Units	**20130712-W01**	**20130712-W02**	**20130712-W03**	**20130712-SG01**	**20130712-SG02**	**20130712-HC01**	**20130712-HC02**	**20130712-HC03**	**20130712-T01**	**20130712-T02**	**20130712-T03**
Microbiological												
Background	CFU/100mL	>2000	>2000	>2000	>2000	>2000	>2000	>2000	>2000	400	480	
Total Coliforms	CFU/100mL	>2000	>2000	>2000	>2000	>2000	>2000	>2000	>2000	60	90	
Escherichia coli	CFU/100mL	200 (1)	60 (1)	70 (1)	360 (1)	370 (1)	80 (1)	60 (1)	90 (1)	20	19	

RDL = Reportable Detection Limit

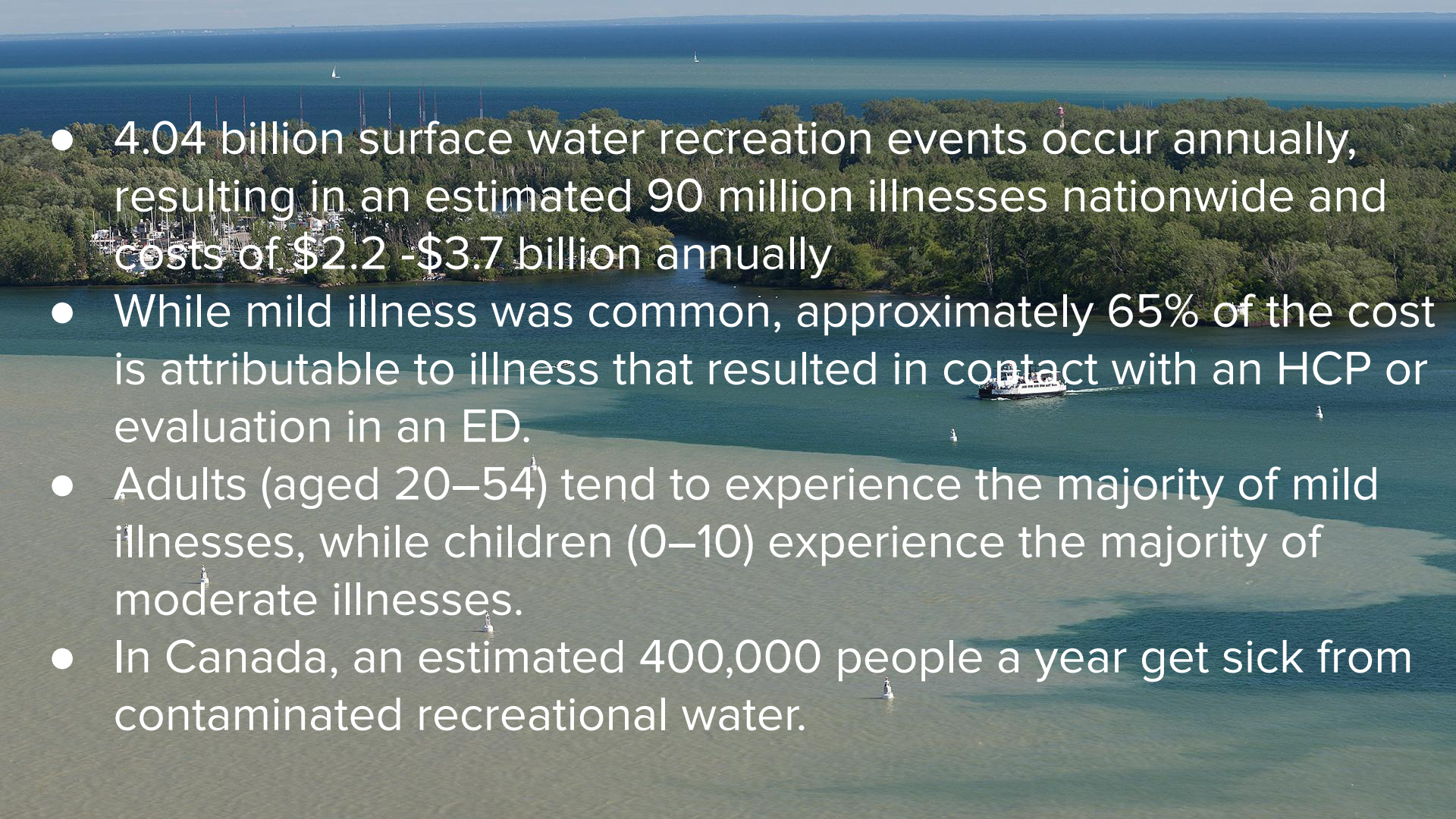
EDL = Estimated Detection Limit

QC Batch = Quality Control Batch

(1) Values reported may be biased low due to overgrowth.

Results relate only to the items tested.


“for the purpose of protecting human health in coastal recreation waters.” US EPA
“protection of public health and safety.” Health Canada”



- 4.04 billion surface water recreation events occur annually, resulting in an estimated 90 million illnesses nationwide and costs of $2.2 -$3.7 billion annually
- While mild illness was common, approximately 65% of the cost is attributable to illness that resulted in contact with an HCP or evaluation in an ED.
- Adults (aged 20–54) tend to experience the majority of mild illnesses, while children (0–10) experience the majority of moderate illnesses.
- In Canada, an estimated 400,000 people a year get sick from contaminated recreational water.

Thanks

gabrielle@swimdrinkfish.ca

www.theswimguide.org

www.recreationalwater.ca

github.com/swimdrinkfish/opendata